Am. J. Hum. Genet. 53:591–608, 1993

mtDNA Variation of Aboriginal Siberians Reveals Distinct Genetic Affinities with Native Americans

Antonio Torroni,* Rem I. Sukernik,‡ Theodore G. Schurr,*,† Yelena B. Starikovskaya,‡ Margaret F. Cabell,* Michael H. Crawford,§ Anthony G. Comuzzie,§ and Douglas C. Wallace*,†

Departments of *Genetics and Molecular Medicine and †Anthropology, Emory University, Atlanta; ‡Section of Human Molecular Genetics, Institute of Cytology and Genetics, Russian Academy of Sciences, Siberian Branch, Novosibirsk; §Department of Anthropology, Laboratory of Biological Anthropology, University of Kansas, Lawrence

Summary

The mtDNA variation of 411 individuals from 10 aboriginal Siberian populations was analyzed in an effort to delineate the relationships between Siberian and Native American populations. All mtDNAs were characterized by PCR amplification and restriction analysis, and a subset of them was characterized by control region sequencing. The resulting data were then compiled with previous mtDNA data from Native Americans and Asians and were used for phylogenetic analyses and sequence divergence estimations. Aboriginal Siberian populations exhibited mtDNAs from three (A, C, and D) of the four haplogroups observed in Native Americans. However, none of the Siberian populations showed mtDNAs from the fourth haplogroup, group B. The presence of group B deletion haplotypes in East Asian and Native American populations but their absence in Siberians raises the possibility that haplogroup B could represent a migratory event distinct from the one(s) which brought group A, C, and D mtDNAs to the Americas. Our findings support the hypothesis that the first humans to move from Siberia to the Americas carried with them a limited number of founding mtDNAs and that the initial migration occurred between 17,000–34,000 years before present.

Introduction

The genetic relationship between aboriginal Siberian and Native American populations has been the object of numerous studies (Rychkov and Sheremtyeva 1977; Sukernik et al. 1981, 1986*a*, 1986*b*, 1988; Szathmary 1981, 1984; Crawford and Enciso 1982). However, with the exception of a few systems, which have been found to be differentially distributed in Siberia and the New World (Gershowitz and Neel 1978; Schanfield 1980; Sukernik and Osipova 1982; Williams et al. 1985; Matsumoto 1988; Schanfield 1992), the majority of the single nuclear gene systems have not been particularly informative in clarifying this relationship.

By contrast, analysis of human mtDNA variation has proved to be a powerful tool for reconstructing ancient Native American migrations (Wallace et al. 1985; Schurr et al. 1990; Torroni et al. 1992, 1993; Wallace and Torroni 1992), although some of the conclusions generated by these analyses have been criticized in other studies (Chakraborty and Weiss 1991; Ward et al. 1991). In Torroni et al. (1992, 1993), the Amerinds and the Na-Dene were shown to have been founded by two migrations which expanded into the Americas at different times. The earlier Amerind migration carried with it only four Asian haplotypes, which subsequently evolved into Amerind-specific haplogroups A, B, C, and D. The later Na-Dene migration carried only haplogroup A mtDNAs.

Each haplogroup was shown to be defined by a specific set of linked polymorphisms, as follows: haplogroup A by an *Hae*III np 663 site gain; haplogroup B by the 9-bp COII-tRNA^{Lys} intergenic deletion (Cann and Wilson 1983; Wrischnik et al. 1987) and an *Hae*III np 16517 site gain; haplogroup C by a linked *Hinc*II np 13259 site loss and an *Alu*I np 13262 site gain; and haplogroup D by an *Alu*I np 5176 site loss. Haplo-

Received December 22, 1992; final revision received May 10, 1993.

Address for correspondence and reprints: Dr. Douglas C. Wallace, Department of Genetics and Molecular Medicine, Emory University, Atlanta, GA 30322.

0002-9297/93/5303-0003$02.00

groups C and D are almost always associated with *Dde*I np 10394 and *Alu*I np 10397 site gains. On the basis of the extent of sequence divergence that has accumulated in the four Amerind haplogroups, the time of arrival of the first Native American migration has been estimated to be at least 17,000 years before present (YBP).

Because our estimate of the time of the first Native American migration was based on the comparison of Asian and Native American mtDNAs (Ballinger et al. 1992; Torroni et al. 1992, 1993), it was possible that some of the Amerind-specific divergence occurred while the ancestral Amerinds were still in Siberia. This would inflate our estimate of the initial migration time. To address this possibility, we characterized the mtDNA variation of 10 aboriginal Siberian populations. This analysis confirmed that Siberian populations are related to the Amerind migration but revealed no overlap between Siberian and Amerind haplogroup A, C, and D mtDNAs. Hence, it appears that all of the Amerind variation occurred within the Americas and that the first migration to the Americas occurred at least 17,000 YBP.

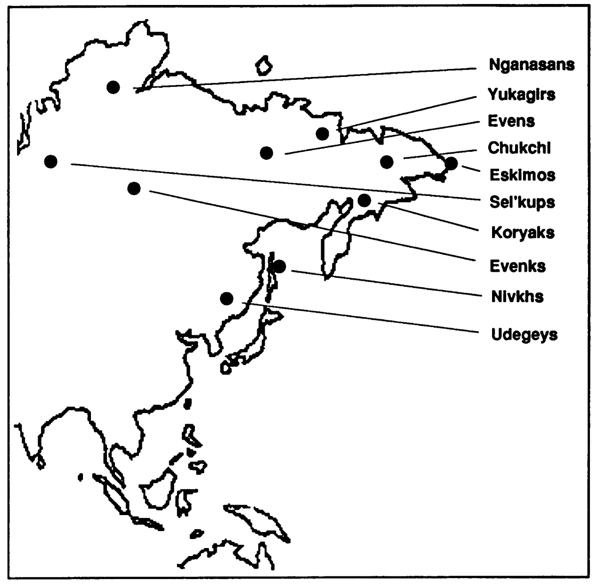


Figure 1 Geographic distribution of the aboriginal Siberian populations analyzed for mtDNA variation.

Subjects and Methods

Populations Sampled

A total of 411 individuals from 10 aboriginal populations of northern Siberia and the Russian Far East were selected for mtDNA characterization. These samples include 20 Sel'kups, 49 Nganasans, 43 Evens, 51 Evenks, 46 Udegeys, 24 Chukchi, 46 Koryaks, 27 Yukagirs, 57 Nivkhs, and 50 Asiatic Eskimos. From each population, a sample of unrelated adults of both sexes was taken from major villages and adjacent camps located in traditional territories. Each village generally numbered approximately 200–250 individuals organized in small consanguineous groups. The Evenk, Nivkh, and Udegey mtDNA samples were extracted from fresh bloods collected between August 1991 and February 1992. The mtDNAs of the remaining groups were extracted from serum samples collected during field expeditions undertaken between 1974 and 1990. The approximate position of the centroids of these tribal samples is shown in figure 1.

The Sel'kups.—The Sel'kups are hunters and fishers living in a few small settlements scattered along the Taz and the Turukhan rivers and their tributaries west of the Lower Yenisey River (Debets 1947; Sukernik et al. 1992). They currently number approximately 1,300 people, and their language belongs to the Samoyedic branch of the Uralic linguistic family. Bloods were taken from individuals living in the village of Farkovo within the easternmost territorial subgroup.

The Nganasans.—The Nganasans are reindeer hunters living in the Taimir Peninsula. Their total population is approximately 700 people, and they speak a language in the Samoyedic branch of the Uralic linguistic family. Although linguistically, culturally, and, to some extent, genetically influenced by Samoyed and Tungus tribes during the past several centuries, the Nganasans apparently derive from northwestern Yukagir tribes which remained unadmixed with surrounding peoples (Karaphet et al. 1981). All three territorial subgroups (Novaya, Volochanka, and Ust-Ayam), subjected by others to conventional population genetic studies (Karaphet et al. 1981; Osipova and Sukernik 1983), are proportionately represented in these samples.

The Evens.—The Evens, also known as Lamuts, are reindeer breeders and fur hunters inhabiting the taiga zone and adjacent boreal forest between the Yenisey River in the west and the Sea of Okhotsk in the east. They currently number around 12,000 people and speak a language in the Tungusic branch of the Altaic linguistic family. They are distributed in scattered aggregates with specific nomadic territories. The Even sera used in this study were collected from individuals living in two geographically separated territories—the upper Yana River (village of Sebyan-Kujhel) and the middle Kolyma River (village of Beryozovka) basins—

with the latter village forming from several families who migrated north from the southern taiga region (Posukh et al. 1990).

The Evenks.—The Evenks are reindeer breeders and fur trappers of the taiga and boreal forest. They number some 20,000 and speak a language in the Tungusic branch of the Altaic linguistic family. They share a common genetic heritage with the Evens, although apparently forming through a mixing of northern aboriginal Siberians and southern populations from the Upper Amur region (Levin and Potapov 1964; Tugolukov 1985). Bloods were taken from Evenk individuals living in the villages of Surinda and Poligus—two adjoining subpopulations located in the middle of the Stony Tunguska River basin—with the majority of the samples analyzed coming from Surinda.

The Udegeys.—The Udegeys are a small, isolated group of aboriginal hunters and fishers who, in recent recorded history, occupied both slopes of the Sikhote-Alin Range south of the Amur River Basin. At the present time, they number no more than 1,000 people. Their original language belongs to the Tungusic branch of the Altaic linguistic family. However, most of them have already been assimilated into the majority of surrounding Russian speakers. Bloods were obtained from unhybridized and unrelated individuals, most of whom were older-age people, in the centralized village of Gvaysugi in the middle of the Khor River basin.

The Chukchi.—At the beginning of the 18th century, the Chukchi numbered approximately 2,000 people, subsisting primarily through hunting game. Since the mid-19th century, they have expanded throughout the northeast region, displacing and partially assimilating Yukagir tribes living to the west and Koryak tribes living to the south of the upper Anadyr River, and experiencing a concomitant population increase to about 10,000 people. They now live in scattered and subdivided groups with small adjoining populations throughout the Chukotka region. Their language belongs to the Chukotko-Kamchatkan linguistic family (Krauss 1988). The sample analyzed represents two local populations in northern Chukotka from two centralized villages, Rytkuchi and Amguyema, and from several adjoining camps.

The Koryaks.—The Koryaks are a hunting, fishing, and reindeer-herding people living in the northern part of the Kamchatka Peninsula. Their total population size is approximately 7,000 individuals, and they speak a language in the Chukotko-Kamchatkan linguistic family (Krauss 1988). Since the end of the Chukchi-Koryak wars in the 19th century, the original Koryak populations have been extensively admixed with reindeer Chukchi, who have expanded southward in territorial range and in population size over the past two centuries (Bogoraz 1910). These samples represent two interrelated populations of northeastern Kamchatka, living in the villages of Achayvayam and Middle Pakhachi and in adjacent reindeer camps (Sukernik et al. 1981).

The Yukagirs.—Until the 13th century, the Yukagirs occupied vast territories of the boreal forest between the Taimir Peninsula and the upper Anadyr River. Since that time, these traditional elk-and-reindeer hunters have been gradually and almost totally assimilated by expanding Even, Yakut, and Chukchi populations (Jochelson 1910; Levin and Potapov 1964; Tugolukov 1979). More recently, the Yukagirs were decimated by epidemics introduced by contact with Russians in the 18th and, particularly, 19th centuries, and they now number fewer than 100 individuals, most of whom are now integrated into other ethnic groups (Levin and Potapov 1964). The relationship of their language with those of the surrounding populations is unclear (Krauss 1988). However, it could be related to the Samoyedic branch of the Uralic linguistic family (Kreynovich 1978). The Yukagir sera consist of two subsamples which equally represent the villages of Nelenmoye and Andrusyhkino in the upper and lower Kolyma River basins, respectively.

The Nivkhs.—Formerly known as the Gilyak, the Nivkhs are hunters and fishers. They consist of two main territorial subdivisions—a continental subgroup dispersed along the Lower Amur River area and a coastal subgroup living mainly along the northwestern and eastern coast of Sakhalin Island. Their current population size is approximately 3,000 individuals. The Nivkh language represents an isolate and is not known to be related to any other (Krauss 1988). Bloods were taken from unrelated and unhybridized individuals living in Rybnovsk and Nekrasovka villages in northern Sakhalin Island.

The Asiatic Eskimos.—The Asiatic Eskimos inhabit the Arctic coast around the Chukotka Peninsula, and their language belongs to the Yupik branch of the Eskimo-Aleut linguistic family. Elder generations show little admixture with the Chukchi or Caucasian populations (Sukernik et al. 1986*b*). Their current overall population size is less than 1,000 people, with almost half of them being admixed with Chukchi and Russians. Most individuals no longer pursue traditional subsistence activities. These Eskimo samples were obtained from unrelated and full-blooded individuals living in, or derived from, three geographically distant communities

—Naukan, Sireniki and Uel'Kal—located around the Chukotka Peninsula.

Molecular Genetic Analysis

mtDNAs were extracted from platelet pellets, buffy coats, and sera, by using the method described by Torroni et al. (1992). Because of the poor quality of the mtDNAs extracted from serum samples, it was impossible to amplify large mtDNA segments by using the PCR (Saiki et al. 1985). This precluded the complete definition of the mtDNA haplotypes for the Asiatic Eskimos, the Chukchi, the Koryaks, the Nivkhs, the Yukagirs, the Nganasans, the Sel'kups, and the Evens. However, it was possible to amplify PCR segments of 200–400 bp from these "suboptimal samples" and to screen them for the presence or absence of the specific mutations which characterize Native American mtDNAs (*Hae*III np 663 site gain, *Alu*I np 5176 site loss, 9-bp intergenic deletion, *Dde*I np 10394 site gain, *Alu*I np 10397 site gain, *Hinc*II np 13259 site loss/*Alu*I np 13262 site gain, *Rsa*I np 16329 site loss, and *Hae*III np 16517 site gain).

For each suboptimal sample, six fragments encompassing the characteristic mutations were PCR amplified using the oligonucleotide primers and conditions reported by Torroni et al. (1993). After endonuclease digestion, the resulting restriction fragments were resolved through electrophoresis in NuSieve plus SeaKem agarose (FMC BioProducts) gels and were visualized by UV-induced fluorescence. This screening generated partial haplotypes on the basis of the presence or absence of the eight polymorphic restriction sites (table 1).

Good-quality DNAs were recovered from Evenk, Udegey, and Nivkh blood samples. This recovery permitted the amplification of the entire mtDNA in nine overlapping segments, by using the primer pairs and amplification conditions described in table 2. Each PCR segment was digested with 14 restriction endonucleases (*Alu*I, *Ava*II, *Bam*HI, *Dde*I, *Hae*II, *Hae*III, *Hha*I, *Hinc*II, *Hin*fI, *Hpa*I, *Msp*I, *Mbo*I, *Rsa*I, and *Taq*I), and the resulting fragments were resolved as described above. This high-resolution analysis generated complete haplotypes for each mtDNA (table 3 and Appendix [table A1]). Table 4 summarizes the mtDNA haplogroup distribution in all 10 Siberian populations.

In addition, 341 bp (np 16030–16370) of the noncoding D-loop region of 11 Evenk, 3 Udegey, and 2 Nivkh mtDNAs were sequenced by using the dideoxy method (Shoffner et al. 1990; Brown et al. 1992) and the primers indicated by Torroni et al. (1993). The mtDNAs to be sequenced were selected on the basis of their haplogroup affiliation (table 5).

Phylogenetic Analysis

The evolutionary relationships of the 34 complete haplotypes observed in the Evenks, Udegeys, and Nivkhs were inferred by using parsimony analysis (PAUP, version 3.0s; Swofford 1992) and were compared with those obtained from Native American populations (Torroni et al. 1992, 1993) (fig. 2). The relationships between Siberian and East Asian mtDNA haplotypes (Ballinger et al. 1992) were delineated by the same method (fig. 3). All dendrograms were rooted by using a Senegalese mtDNA haplotype ("African outgroup"; Appendix [table A1]). For each analysis, maximum parsimony (MP) trees were generated through random addition of haplotypes, by using both the Tree Bisection and Reconnection (TBR) and Nearest-Neighbor Interchange (NNI) branching algorithms. Because of the large number of terminal taxa, a very large number of MP trees were obtained with both branch-swapping methods. We terminated our searches at 3,000 MP trees and saved no more than 10 MP trees for each replication. Consequently, shorter trees could exist, although none was observed in these analyses. Strict consensus trees of the 3,000 MP trees generated by both methods were also obtained. Consensus trees are hierarchical summaries of the information common to a set of MP trees. A strict consensus tree contains only those groups appearing in all MP trees (Sokal and Rohlf 1981). However, even when only 3,000 of the MP trees were used to generate a strict consensus tree, most of the relationships between haplotypes were unresolved.

The Siberian D-loop sequences were aggregated with comparable data from Native American and East Asian populations (Torroni et al. 1993) and were subjected to parsimony analysis (fig. 4). The dendrograms were generated with the TBR branching algorithm, by means of random addition of sequences, and were rooted from a !Kung D-loop sequence (subject 1; Vigilant et al. 1989). As for the haplotype trees, although no shorter trees were found, they could exist, and a large number of MP trees were obtained in each search.

Sequence Divergences

Intragroup sequence divergences from restriction analysis data were estimated with the maximum likelihood procedure of Nei and Tajima (1983) by using the program DREST (provided by L. Jin). The methodology of this program is given by Torroni et al. (1992). When the divergence times are calculated for mtDNA haplogroups shared between aboriginal Siberians and Native Americans, a mtDNA evolution rate of 2.0%–

Table 1

Distribution of Partial mtDNA Haplotypes in Aboriginal Siberians

	Population						
Partial Haplotype[a]	Eskimos (*N* = 50)	Chukchi (*N* = 24)	Koryaks (*N* = 46)	Evens (*N* = 43)	Yukagirs (*N* = 27)	Nganasans (*N* = 49)	Sel'kups (*N* = 20)
Haplogroup A:							
+663e	40	8	10	0	0	1	0
+663e, +16517e	0	1	1	0	0	0	0
Total	40	9	11	0	0	1	0
Haplogroup C:							
−13259o, +13262a, +16517e	0	2	6	0	0	0	0
−13259o, +13262a, +10394c, +10397a, +16517e	0	2	4	25	16[b]	19	7
Total	0	4	10	25	16	19	7
Haplogroup D:							
−5176a	2	0	3	0	7	0	0
−5176a, +10394c, +10397a	7	4	1	3	2	8	0
−5176a, +10394c, +10397a, +16517e	1	0	0	0	0	10	0
Total	10	4	4	3	9	18	0
Others[c]:							
+16517e	0	5	13	0	0	7	8
+10394c, +10397a	0	0	0	0	0	2	0
+10394c, +10397a, +16517e	0	2	7	15	2	1	0
No mutations	0	0	1	0	0	1	5
Total	0	7	21	15	2	11	13

Note.—mtDNAs extracted from Siberian serum samples were PCR amplified as six small fragments and were screened for the eight genetic markers which define mtDNA haplogroups A–D. The observed combinations of restriction sites are listed for haplogroups A, C, and D, together with the no. of individuals from each population who harbor that haplotype.

[a] Single letters are used to identify each restriction enzyme, as follows: a = *Alu*I; c = *Dde*I; e = *Hae*III; and o = *Hinc*II.

[b] Three Yukagir mtDNAs also have the 9-bp COII-tRNA[Lys] intergenic deletion.

[c] Indicates mtDNAs lacking the polymorphisms which define haplogroups A–D.

4.0%/million years (MYR) was used (Stoneking et al. 1986; Cann et al. 1987; Wallace et al. 1987). This standard rate of human mtDNA evolution does not incorporate any variance components. Therefore, a 95% confidence interval for the rate would be considerably broader.

Results

Haplotype Analysis

To investigate the relationships between Siberian and Native American mtDNAs, 259 aboriginal Siberian mtDNAs from seven populations were screened for the

Table 2

PCR Primer Pairs Used for Amplification of the Entire mtDNA in Nine Fragments

Primer Coordinates[a]	T_H[b] (°C)	Size (bp)
16453–16472,[c] 1696–1677[d]	61	1,812
1562–1581,[e] 3717–3701[f]	51	2,155
3108–3127, 5917–5898[g]	53	2,809
5591–5610, 7433–7414	53	1,842
7367–7384, 9172–9154	51	1,796
8282–8305, 10107–10088[h]	55	1,825
9802–9821, 11873–11851[i]	59	2,071
11673–11691, 13950–13932[j]	57	2,277
13809–13828, 16547–16527[k]	57	2,738

[a] Numbered according to Anderson et al. (1981).
[b] Calculated from the primer sequence without the tail sequence added.
[c] Has 5′-CCACCCTGCAG tail.
[d] Has 5′-CCACAAGCTT tail.
[e] Has 5′-CCACCTGCAG tail.
[f] Has 5′-CCACAAGCTT tail.
[g] Has 5′-CCACAAGCTT tail.
[h] Has 5′-CCACAAGCTT tail.
[i] Has 5′-CCAAGCTTCCA tail.
[j] Has 5′-CCAAGCTT tail.
[k] Has 5′-CCACAAGCTT tail.

markers which define Native American mtDNA haplogroups A–D (table 1). Additionally, 57 Nivkh, 51 Evenk, and 45 Udegey mtDNAs were characterized by high-resolution restriction analysis to generate complete mtDNA haplotypes (table 3). The high-resolution mapping revealed 34 distinct haplotypes defined by (*a*) 58 polymorphic sites and (*b*) the 4-bp COII-tRNA^{Lys} intergenic insertion (Cann and Wilson 1983; Ballinger et al. 1992) (table 3 and Appendix). Among these haplotypes, two (S18, S19) were defined by the mutations specific to haplogroup A, nine (S26–S34) by the those specific to haplogroup C, and seven (S10–S17) by those specific to haplogroup D.

These restriction analyses revealed that mtDNAs from only three of the four haplogroups (A, C, and D) observed in Native Americans were found in Siberians. None of the Siberian mtDNAs belonged to the fourth haplogroup, group B, and the Asiatic Eskimos showed only haplogroup A and D mtDNAs (tables 1 and 3). The frequencies of each mtDNA haplogroup in the 10 aboriginal Siberian populations are presented in table 4.

Haplogroup A haplotypes represented 15.3% of the Siberian mtDNAs. Haplotypes from this group have been previously observed in virtually all Northern Na-Dene and in 34.4% of the Amerinds (Torroni et al. 1993). They have also been described in 7.5% of the Han Chinese and 7.7% of the Koreans (Ballinger et al. 1992). In Siberia, haplogroup A mtDNAs appeared in a distinct, geographically limited distribution (table 4). They were observed at their highest frequencies in the northeasternmost Siberian populations, representing 23.9% of the Koryak, 37.5% of the Chukchi, and 80.0% of the Asiatic Eskimo samples. These mtDNAs were absent or at low frequencies in the western (Sel'kups, Evenks, Evens, Nganasans, and Yukagirs) and southern (Nivkhs and Udegeys) tribes.

While all Siberian haplogroup A mtDNAs harbored the characteristic *Hae*III np 663 site gain, only 3.2% also had the *Hae*III np 16517 site gain (tables 1 and 3). The *Hae*III np 16517 site is located in the D-loop region and is hypervariable. Therefore, this site is present in a portion of the Native American haplotypes within each of the haplogroups A, C, and D (Torroni et al. 1993). The absence of this site in most Siberian haplogroup A mtDNAs suggests that the founding Native American haplogroup A mtDNA had only the *Hae*III np 663 site (AM1; Torroni et al. 1993).

Haplogroup C mtDNAs were found to be common in aboriginal Siberians, representing 32.4% of their mtDNAs. Moreover, these haplotypes were present in 8 of the 10 Siberian populations analyzed and were found in 84.3% of the Evenks (tables 3 and 4). In all cases, the mutations that define haplogroup C (*Hin*cII np 13259 site loss; *Dde*I np 10394 and *Alu*I np 10397 site gains) were associated with the *Hae*III np 16517 site gain (tables 1 and 3). This implies that the founding haplogroup C mtDNA for the Amerinds had all four of these mutations, including the *Hae*III np 16517 site (AM43; Torroni et al. 1993). Because 19.6% of the Amerinds (Torroni et al. 1993), 1.6% of "Orientals" (Blanc et al. 1983), 1.8% of the Japanese (Horai et al. 1984), and 2.9% of the Han Chinese (Ballinger et al. 1992) harbor haplogroup C mtDNAs, this finding firmly links Native American and Siberian populations.

Haplogroup D mtDNAs were also observed in 8 of the 10 Siberian populations analyzed, with an overall frequency of 16.8%. The highest haplogroup D frequencies occurred in the Nganasans (36.7%) and the Yukagirs (33.3%). mtDNAs belonging to this haplogroup were previously described in 19.3% of the Amerinds (Torroni et al. 1993), as well as in several East and Southeast Asian populations (Ballinger et al. 1992). In addition to the *Alu*I np 5176 site loss and the *Dde*I np 10394 and *Alu*I np 10397 site gains, the majority of the

Table 3

Distribution of Complete mtDNA Haplotypes in Aboriginal Siberians

		Population			
Haplotype[a]	Haplogroup	Nivkhs	Evenks	Udegeys	*N*
S1	Other[b]	26	...	2	28
S2	Other	1	...	...	1
S3	Other	4	...	...	4
S4	Other	5	...	...	5
S5	Other	1	...	...	1
S6	Other	...	...	1	1
S7	Other	...	...	1	1
S8	Other	2	...	...	2
S9	Other	1	...	...	1
S10	D	13	...	...	13
S11	D	1	...	...	1
S12	D	2	...	...	2
S13	D	...	2	...	2
S14	D	...	1	...	1
S15	D	...	1	...	1
S16	D	...	1	...	1
S17	Other	1	...	...	1
S18	A	...	1	...	1
S19	A	...	1	...	1
S20	Other	...	1	...	1
S21	Other	...	...	8	8
S22	Other	...	...	1	1
S23	Other	...	...	7	7
S24	Other	...	...	13	13
S25	Other	...	...	4	4
S26	C	...	11	...	11
S27	C	...	6	7	13
S28	C	...	7	...	7
S29	C	...	10	...	10
S30	C	...	2	1	3
S31	C	...	2	...	2
S32	C	...	2	...	2
S33	C	...	2	...	2
S34	C	...	1	...	1
Total		57	51	45	153

[a] Polymorphic restriction sites defining the 34 haplotypes observed in Siberia are listed in the Appendix.
[b] "Other" haplogroups comprise those haplotypes not belonging to haplogroups A, B, C, and D.

Siberian group D mtDNAs lacked the *Hae*III np 16517 site gain (tables 1 and 3), suggesting that the founding Amerind group D haplotype lacked the *Hae*III site gain (AM88; Torroni et al. 1993).

Haplogroup B mtDNAs were absent in Siberian populations (tables 1 and 3), in marked contrast with the Amerinds and East Asians in which they are common (Horai and Matsunaga 1986; Ballinger et al. 1992; Harihara et al. 1992; Torroni et al. 1993). However, 3 of the 27 Yukagir mtDNAs did have the 9-bp deletion. As these were associated with the three mutations which define haplogroup C (table 1), this finding indicates that the 9-bp deletion arose independently in the Yukagirs in a haplogroup C mtDNA. Independent deletion events have also been found in other Asian and Native American mtDNAs (Schurr et al. 1990; Ballinger et al. 1992; Torroni et al. 1993).

With the exception of the Asiatic Eskimos, all of the Siberian populations examined harbored significant numbers of mtDNAs which did not fall into haplogroups A–D. These "other" mtDNAs (tables 1 and 3) varied in frequency from 2.0% in the Evenks to 82.6%

Table 4

Percent Frequencies of mtDNA Haplogroups in Aboriginal Siberians

Population	Haplogroup[a] A	B	C	D	Other	N
Eskimos	80.0	...	...	20.0	...	50
Chukchi	37.5	...	16.7	16.7	29.2	24
Koryaks	23.9	...	21.7	8.7	45.6	46
Yukagirs	...	...	59.3	33.3	7.4	27
Evens	...	...	58.1	7.0	34.9	43
Nivkhs	...	...	...	28.1	71.9	57
Udegeys	...	...	19.6	...	80.4	46
Evenks	3.9	...	84.3	9.8	2.0	51
Nganasans	2.0	...	38.8	36.7	22.4	49
Sel'kups	...	...	35.0	...	65.0	20

[a] The haplotypes grouped into haplogroups A, B, C, D, and Other are expressed as a percentage of the total no. of individuals in the tribe who were analyzed.

in the Udegeys. About 63% of these mtDNAs were characterized by the *Dde*I np 10394 site gain, with a significant proportion of these, about 56%, also having the *Alu*I np 10397 site gain. Previous studies have shown that mtDNAs with and without the *Dde*I np 10394 and *Alu*I np 10397 sites define two subgroups of mtDNAs common in East and Southeast Asian populations (Ballinger et al. 1992).

The Siberian mtDNAs were also screened for the absence of the *Rsa*I np 16329 site, a mutation found in 29.0% of the Na-Dene haplogroup A mtDNAs but not in those of Amerinds (Torroni et al. 1992, 1993). This marker was not observed in any of the Siberian mtDNAs analyzed in the present study (tables 1 and 3), nor in the Alaskan and Siberian Eskimo, Aleut, and Chukchi mtDNAs analyzed by Shields et al. (1992). Consequently, this mutation appears to have arisen in the Americas after the ancestral Na-Dene separated from the modern aboriginal Siberians and Amerinds (Torroni et al. 1992, 1993).

Phylogenetic Analysis

The genetic affinities between the aboriginal Siberian and Native American mtDNAs were further defined by means of parsimony analysis. One phylogeny encompassing 34 Siberian (S1–S34) and 92 Native American (AM1–AM63, AM65–AM70, and AM74–AM96) haplotypes is shown in figure 2. This dendrogram is one of the thousands of MP trees generated by the TBR branch-swapping algorithm. It is 182 mutational steps in length, with consistency index (CI) and retention index (RI) of .576 and .895, respectively. Figure 2 shows one of the MP trees in which *all* Siberian and Native American haplotypes belonging to haplogroups A, C, and D segregated together into the corresponding haplogroup branches. Group B included only Native American haplotypes. The "other" Siberian haplotypes were positioned outside haplogroups A–D, with some of them interspersed with Native American haplotypes AM28, AM29, and AM74–AM76, which are thought to be of European origin (Torroni et al. 1993).

A comparatively large number of MP trees were obtained in the analysis of the Siberian haplotypes and 106 Asian haplotypes (AS17–AS122; Ballinger et al. 1992). One of these trees is reported in figure 3. It is 273 mutational steps in length and has CI and RI of .478 and .798, respectively (for a definition of these indices, see the accompanying article [Torroni et al. 1993]). This tree is one in which almost all Siberian and Asian haplotypes characterized by the specific polymorphisms of haplogroups A, B, C, and D cluster together. In this case, the exception is represented by group D haplotype S16, which fell outside cluster D (fig. 3). This discrepancy is probably due to a secondary mutation which simultaneously eliminated both the *Dde*I np 10394 and the *Alu*I np 10397 site gains characteristic of haplogroup D haplotypes. All of the "other" haplotypes were scattered throughout the tree, usually in association with other Asian mtDNAs.

Strict consensus trees of subsets of MP trees were also generated for both the Siberian–Native American and the Siberian–Asian parsimony analyses. However, even when 3,000 of the MP trees were used to generate them, the relationships between most haplotypes were unresolved (figs. 2 and 3; insets). This further indicated that parsimony analysis is unable to resolve the deep branches of trees when the number of mtDNA haplotypes and character states is large (Hedges et al. 1991; Templeton 1991).

D-Loop Sequence Analysis

The relationship between Siberian, Asian, and Native American mtDNAs was also examined by comparing 341 bp of the D-loop sequence from representatives of haplogroups A, C, and D. Direct DNA sequencing revealed 23 variable nucleotide positions in the 16 Siberian samples studied (table 5). Two Evenk group A mtDNAs were sequenced. While they were associated with different mtDNA haplotypes (S18 and S19), they had identical D-loop sequences (50 and 51). These sequences exhibited mutations at np 16290 (T to C) and

Table 5

D-Loop Sequences from Aboriginal Siberians

Sample	Haplotype	Polymorphic Nucleotide Positions[a]
		11111111111111111111111111111111111111111111111111111111111
		66666666666666666666666666666666666666666666666666666666666
		00000001111111111111112222222222222222222222222333333333333
	Haplo-	34558999012224455777899011123334446667778999999011122334566
	type	92196235414690804124928324732350890361488013458014957154602
Cambridge[b]		GGAATTTCCCTTGTCCTATCTCTAACTCCAAACTCTCTGCTCCACCTATAGTCAACTCT
Haplogroup A:		
50 Evenk	S18	A-------------------C------T-------------**T**--------**A**-----C-C
51 Evenk	S19	A-------------------C------T-------------**T**--------**A**-----C-C
Haplogroup C:		
52 Evenk	S27	--**C**-----**T**------
53 Evenk	S26	-----------------G---------T------------------**C**-----**T**------
54 Evenk	S28	------C--------------------T------------C-----**C**-----**T**------
55 Evenk	S29	---------------------------T------------------**C**-----**T**------
56 Evenk	S32	---------------------------T-----C--------T--T**C**-----**T**------
57 Udegey	S27	---------------------------T------------------**C**-----**T**------
58 Udegey	S27	-------T-------------------T------------------**C**-----**T**------
59 Udegey	S30	------------A--T-----------T------------------**C**-----**T**------
Haplogroup D:		
60 Nivkh	S10	-A-----------------------T-T------------------------------C
61 Nivkh	S12	------C--------------------TT------------T----------------C
62 Evenk	S14	---------------------------T------------------------------C
63 Evenk	S16	-----C---------------------T--------T---------------------C
64 Evenk	S15	---------------------------T----------------------A-------C
65 Evenk	S13	---------------------------T--------------T---------------C

[a] Boxes indicate the nucleotide variants specific to each D-loop group (Torroni et al. 1993).
[b] From Anderson et al. (1981).

np 16319 (G to A) specific to D-loop group A, which were also observed in the Koreans and the Han Chinese (Torroni et al. 1993). Neither the Evenk, Korean, nor Han sequences had the C-to-T transition at np 16111, which is characteristic of both the Na-Dene and Amerind D-loop group A sequences (Torroni et al. 1993). This finding suggests that Siberia was colonized by mtDNAs lacking the np 16111 mutation, and that this mutation subsequently arose in a Siberian subpopulation which was ancestral to both the Amerinds and the Na-Dene.

The D-loop sequences of eight Siberian haplogroup C mtDNAs were also determined. All of them showed the D-loop group-specific polymorphisms at np 16298 (T to C) and np 16327 (C to T) (Torroni et al. 1993). Two of these, one Evenk (55) and one Udegey (57), had D-loop sequences identical to that observed in one Han Chinese from Tibet (39) and represent the "consensus" of all D-loop group C sequences observed in Asia and Siberia. Consequently, this sequence may be the founding D-loop group C sequence in Asia and Siberia (table 5). Neither the Han Chinese nor Siberian mtDNAs had the T-to-C transition at np 16325 which is found in most of the Amerind D-loop group C sequences (Torroni et al. 1993). Therefore, this mutation probably occurred in the Siberian population which subsequently founded the Amerinds.

Six D-loop group D sequences from two Nivkhs and four Evenks were determined. Two transitions, a C to T at np 16223 and a T to C at np 16362, were observed in all of these sequences. However, these two polymorphisms were also found in D-loop group A and C sequences. In addition, Evenk mtDNA 62 showed a D-loop sequence identical to that reported in one

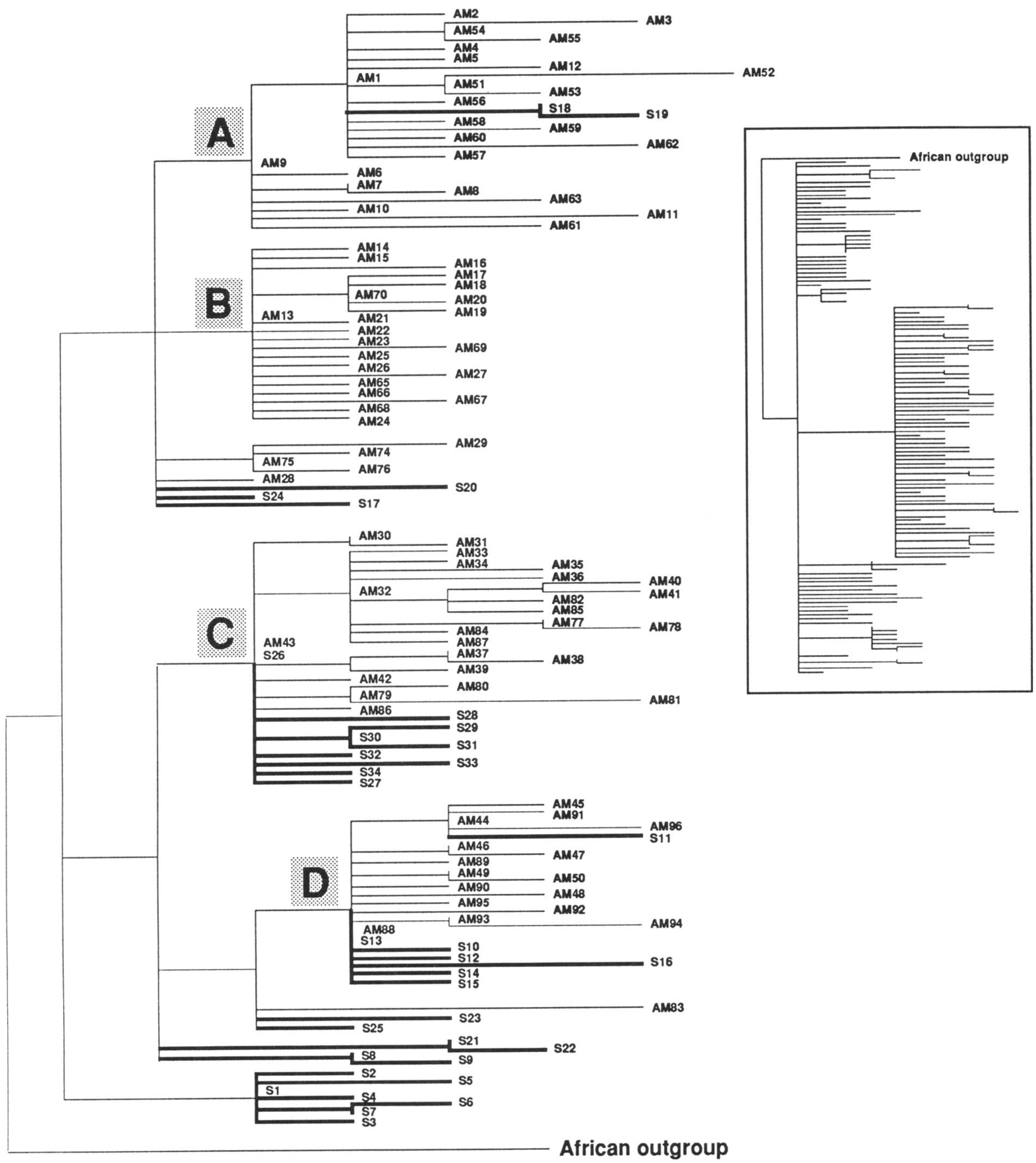


Figure 2 Phylogenetic tree of 34 Siberian (S1–S34) and 92 Native American (AM1–AM63, AM65–AM70, and AM74–AM96) haplotypes. Native American data are from Torroni et al. (1992, 1993). This dendrogram is 182 steps in length and has CI and RI of .576 and .895, respectively. It is one of thousands of MP trees generated by the TBR method. Thickened lines indicate Siberian haplotypes. The letters in shaded boxes, A, B, C, and D, indicate the four mtDNA haplogroups. The number at the end of each branch represents a unique mtDNA haplotype. The horizontal branch lengths are proportional to the number of mutational events between haplotypes. The inset shows the strict consensus of 3,000 MP trees generated with the TBR algorithm. The length of the consensus tree is 308 steps, with CI and RI of .232 and .528, respectively.

Taiwanese Han (Torroni et al. 1993), the consensus for all D-loop group D sequences found in Asia, Siberia, and the Americas (table 5). Therefore, it is possible that the D-loop sequence 62 founded D-loop group D in both Siberia and the Americas.

A phylogeny including the Siberian, Native American, and Asian D-loop sequences (Torroni et al. 1993) revealed a complete correspondence between haplotype and D-loop groups A, B, and C (fig. 4). However, the D-loop group D sequences did not cluster together because of the lack of group-specific polymorphisms. Moreover, because of the high evolution rate of the D-loop, parallel nucleotide substitutions are relatively common (table 5). Examples are the C residue at np 16093 (54 and 61), the T residue at np 16290 (50 and 61), the T residue at np 16291 (56 and 65), and the A residue at np 16319 (50 and 64). In the absence of group-specific polymorphisms, such parallel mutations strongly influence the structure of the phylogeny (fig. 4). Therefore, as observed in Native American mtDNA studies (Torroni et al. 1993), Siberian mtDNAs are more clearly defined by high-resolution restriction analysis than by sequencing of only one of the two hypervariable D-loop sequences (Stoneking et al. 1992). Most important, contrary to the conclusions of Ward et al. (1991), these D-loop–sequence studies support the conclusion drawn from mtDNA haplotype analysis that Native American populations were founded by a limited number of mtDNAs. The haplotype and D-loop–sequence data extend this conclusion by indicating that aboriginal Siberian populations were also founded by a limited number of mtDNAs.

Discussion

The analysis of mtDNA variation of 411 Siberians from 10 aboriginal populations indicates that Siberians derived from Asians and that, in turn, Native Americans derived from Siberians. Moreover, Siberians were found not only to harbor three of four mtDNA haplogroups (A, C, and D) found in Native Americans but also to contain a significant proportion of other mtDNAs of probable Asian origin. This suggests that both Siberian and Native American populations derived from Asian populations which underwent a series of population constrictions. Finally, within the common haplogroups A, C, and D, only the nodal haplotypes are shared between Siberians and Native Americans. This result indicates that most of the Siberian and Native American mtDNA variation accumulated after ancestral Americans entered the New World.

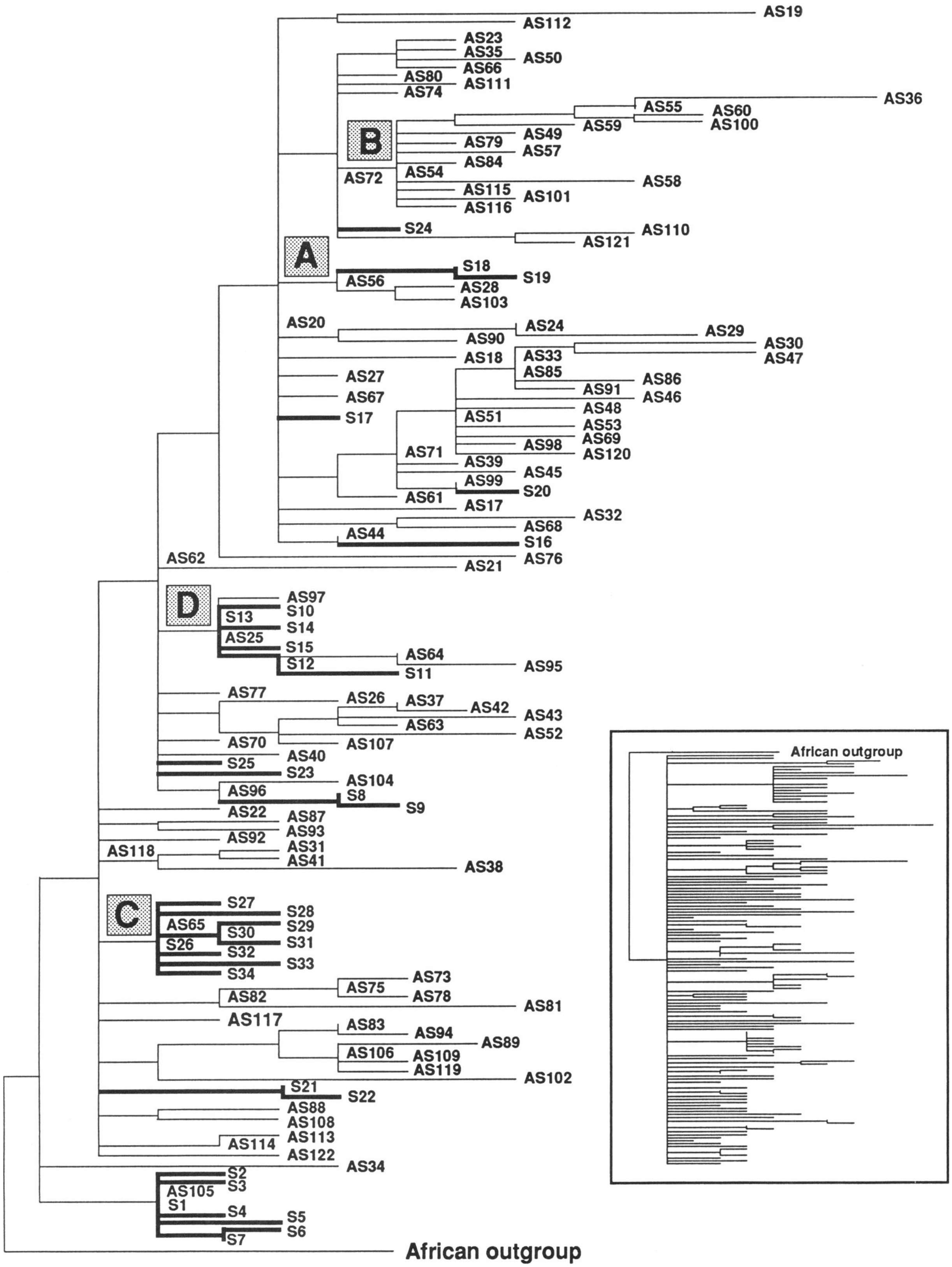
Haplotypes Shared between Siberians and Native Americans

Only two haplotypes were shared between Siberians and Native Americans: group C haplotypes S26 and AM43 and group D haplotypes S13 and AM88. Haplotype S26 was observed in 21.6% of the Evenks and in 33.3% of the Siberian group C haplotypes and formed the node of Siberian and Native American haplogroup C (fig. 2). Therefore, this haplotype was probably the founding mtDNA of group C in Siberia and the only group C haplotype which moved from Siberia to the Americas. Similarly, haplotype S13 was observed in 28.1% of the Nivkhs and in 13.7% of the overall Siberian group D haplotypes and represented the node of Siberian and Native American group D haplotypes (fig. 2). Thus, it is probably the founding haplotype of haplogroup D for both geographic regions.

The Affinity between Asian and Siberian mtDNA Haplotypes

The origins of Siberian populations were further delineated by comparing complete Siberian haplotypes to those from East Asians (Ballinger et al. 1992). The clear relationship of the Asian populations to Siberians is illustrated by the dendrogram of figure 3. Noteworthy associations include the linked *Hin*cII/*Hpa*I np 12406 site losses found in Evenk haplotype S20 which has also been observed in Koreans and Han (AS98, AS99, and AS61; Ballinger et al. 1992), Japanese (Horai et al. 1984), and several other Asian populations (Brega et al. 1986; Cann et al. 1987; Harihara et al. 1988; Ballinger et al. 1992) and which may represent one of the earlier Mongoloid mtDNA lineages (Blanc et al. 1983; Ballinger et al. 1992). Similarly, Nivkh haplotypes S8 and S9 have features similar to others observed in Koreans (AS96 and AS104; Ballinger et al. 1992) and Japanese (Horai et al. 1984; Horai and Matsunaga 1986), while Nivkh and Udegey haplotypes S1–S7 define a distinct haplogroup in both dendrograms (figs. 2 and 3) whose nodal haplotype, S1, is very similar to haplotypes observed by others in Japanese (Horai and Matsunaga 1986) and is identical to the haplotype AS105 observed in Koreans (Ballinger et al. 1992). The two Siberian haplotypes, S13 and S26, shared with the native Americans have identical Asian counterparts, AS25 observed in Koreans and Han Chinese and AS65 described in the Taiwanese Han (Ballinger et al. 1992). Therefore, the Siberians of the Amur region clearly show a close genetic affinity with the Japanese, the Koreans, and the Han.

The origin of Siberian mtDNA haplogroups A, C,


AS19
AS112
AS23
AS35
AS66
AS50
AS80
AS111
AS74
AS36
AS55
AS60
AS59
AS100
B
AS49
AS79
AS57
AS84
AS54
AS58
AS72
AS115
AS101
AS116
S24
AS110
AS121
A
S18
S19
AS56
AS28
AS103
AS20
AS24
AS29
AS90
AS30
AS47
AS18
AS33
AS85
AS86
AS27
AS91
AS67
AS46
AS48
S17
AS51
AS53
AS69
AS98
AS71
AS120
AS39
AS45
AS99
S20
AS61
AS17
AS32
AS68
AS44
S16
AS62
AS76
AS21
D
AS97
S13
S10
S14
AS25
S15
AS64
S12
AS95
S11
AS77
AS26
AS37
AS42
AS43
AS63
AS52
AS70
AS107
S25
AS40
S23
AS104
AS96
S8
S9
AS22
AS87
AS93
AS92
AS31
AS118
AS41
AS38
C
S27
S28
AS65
S30
S29
S26
S32
S31
S33
S34
AS73
AS75
AS82
AS78
AS81
AS117
AS83
AS94
AS89
AS106
AS109
AS119
AS102
S21
S22
AS88
AS108
AS114
AS113
AS122
AS34
S2
S3
AS105
S1
S4
S5
S6
S7
African outgroup
African outgroup

and D in Asia can be defined by comparing the Asian, Siberian, and Native American mtDNA haplotypes, D-loop sequences, and phylogenies. The common group C haplotype from the three population groups is designated S26/AM43/AS65, the common group D haplotype is S13/AM88/AS25, and the common group A haplotype is AM1/AS56 (Torroni et al. 1993). The D-loop sequences support these associations and reveal that population movements carried these mtDNAs from central east Asia to Siberia and then to the Americas. For example, the D-loop group A sequence associated with the Taiwanese Han haplotype AS56 has a C residue at np 16114, while Native American haplotype AM1 (identical to AS56) and virtually all of its Native American derivatives have a T residue at the same position. The Evenk group A mtDNAs have a C residue, as observed in the Taiwanese Han. Similarly, for haplogroup C, the Han haplotype AS65 and the Siberian haplotype S26 have a T residue at np 16325, while the Native American haplotype AM43 (identical to AS65 and S26) and all its Native American derivatives have a C residue at this same position. The population distribution of these mutations suggests a progression of population movements and indicates that Siberia was colonized by people related to modern Han and Koreans. However, the frequency of group A, C, and D haplotypes in modern Han Chinese and Koreans is low relative to Siberians. Thus, while Siberian mtDNAs clearly derived from Asian mtDNAs, the marked change in mtDNA frequencies in Siberia implies that there was a substantial population constriction in forming the Siberian populations.

While the marked increase in group A, C, and D haplotypes in many Siberian populations implies a substantial population constriction, the constriction was probably not as complete as that which gave rise to Native Americans. This is apparent from the 36.1% of Siberian mtDNAs not belonging to haplogroups A, C, and D. These "other" mtDNAs show clear Asian affinities but are absent in Asiatic Eskimos, Na-Dene, and Amerinds. While they could have derived from recent admixture of Siberian and Asian populations, they are more likely to have been carried to Siberia by the Asian migration and subsequently to have been lost by the Native American migrations.

The sequential reduction in complexity from Asia through Siberia to the Americas implies that each migration was accompanied by genetic bottlenecks. However, the complexities of mtDNA genetics still leave it unclear as to the actual size of the populations involved in the colonizations. For example, the uniparental inheritance of the mtDNA (Giles et al. 1980) greatly reduces the effective population size. Moreover, population expansions will affect the mtDNA diversity remaining in extant populations. Hence, it is quite reasonable that the mtDNA variation of an aboriginal population might be low, while the nuclear genetic variation would remain high (Kidd et al. 1991).

Haplogroup B Distribution in Asia and the Americas

The most striking anomaly in the Siberian data is the absence of group B deletion haplotypes. As shown in figure 5, the deletion haplotypes are widely distributed in East Asians (Cann and Wilson 1983; Horai and Matsunaga 1986; Ballinger et al. 1992; Harihara et al. 1992) and occur at very high frequencies in Melanesian and Polynesian populations (41%–100%; Hertzberg et al. 1989; Stoneking et al. 1990). Moreover, group B mtDNAs are dispersed throughout the Amerinds of North, Central, and South America at continental frequencies of 19%, 33%, and 18%, respectively (Torroni et al. 1993). Yet, the geographic space that connects the Asian and Amerind populations is devoid of this variant. Two alternative possibilities may explain this anomaly. Haplogroup B mtDNAs may have been present in Siberia along with haplogroup A, C, and D mtDNAs but were subsequently lost by drift. Alternatively, group B mtDNAs arrived in the Americas by a different route than that followed by haplogroup A, C, and D mtDNAs. The loss of group B by genetic drift is possible. However, this hypothesis is not entirely satisfactory, because it is unlikely that all 10 populations would lose the same founding mtDNA lineage, and because the genetic diversity of the Amerind group B haplotypes is less than that of haplogroups A, C, and D (Torroni et al. 1993).

As an alternative to genetic drift, the absence of haplogroup B in Siberia could be explained as the product of two separate migrations, the first carrying haplogroup A, C, and D mtDNAs and the second carrying haplogroup B mtDNAs. The most likely route for such

Figure 3 Phylogenetic tree of 34 Siberian (S) and 106 Asian (AS) haplotypes. Asian data are from Ballinger et al. (1992). Thickened lines indicate Siberian haplotypes. Haplogroups correspond to those indicated in fig. 2. This dendrogram is one of thousands of MP trees generated by the TBR method. The inset shows the strict consensus of 3,000 MP trees generated with the TBR algorithm. The length of the consensus tree is 435 steps, with CI and RI of .235 and .399, respectively.

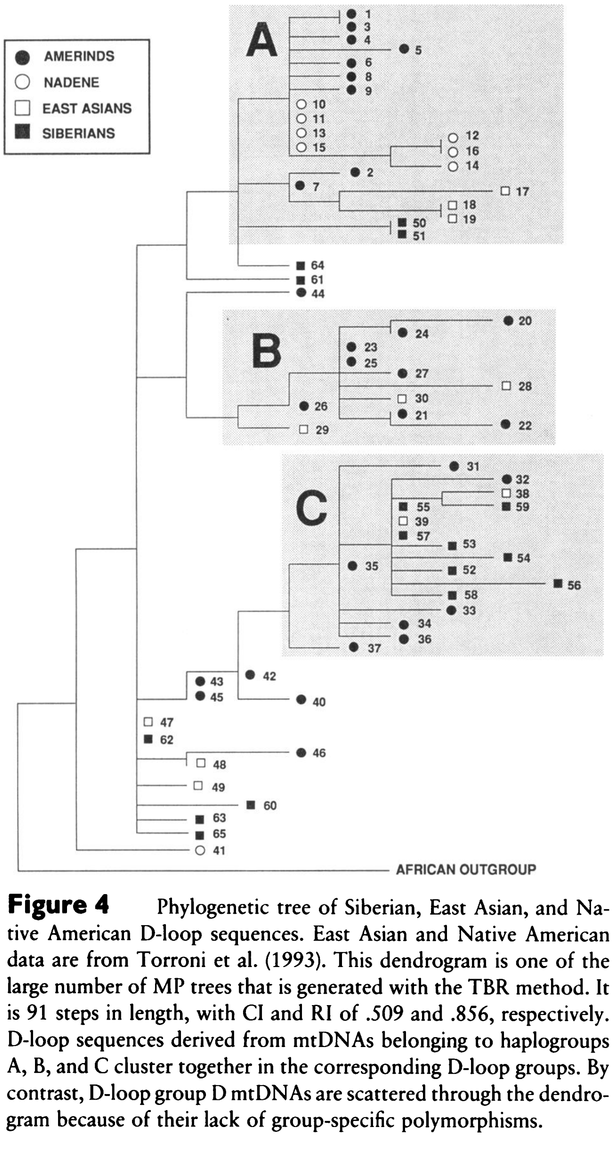


Figure 4 Phylogenetic tree of Siberian, East Asian, and Native American D-loop sequences. East Asian and Native American data are from Torroni et al. (1993). This dendrogram is one of the large number of MP trees that is generated with the TBR method. It is 91 steps in length, with CI and RI of .509 and .856, respectively. D-loop sequences derived from mtDNAs belonging to haplogroups A, B, and C cluster together in the corresponding D-loop groups. By contrast, D-loop group D mtDNAs are scattered through the dendrogram because of their lack of group-specific polymorphisms.

an alternative migration of group B mtDNAs would be an expansion along the coast of Siberia. By this coastal route, the group B migration could have avoided contacts with Asiatic peoples inhabiting the tundra of eastern Siberia. Such a two-migration model could account for the relative paucity of genetic variation within Amerind group B haplotypes relative to those from groups A, C, and D and for the prevalence of the founding haplotypes of haplogroup B (AS54 = AM13) in both Asians and Native Americans (Ballinger et al. 1992; Torroni et al. 1992, 1993). In addition, the sequence divergence value of haplogroup B in the Amerinds was estimated to be only .024%, whereas those for the Amerind haplogroups A, C, and D were 0.091%, 0.096%, and 0.053, respectively (Torroni et al. 1993).

A possible difficulty with this hypothesis is that such a second migration to the Americas might have been expected to encounter resistance from the resident populations. However, this might not have been a problem if the population density in the Americas was very low or if the second migratory group joined the first at the Bering land bridge.

Quantification of Siberian mtDNA Diversity

The fact that the Siberian and Native American mtDNA groups C and D share only the nodal and founding haplotypes (S26/AM43 for group C and S13/AM88 for group D) indicates that most of the mtDNA variation of the populations from these two regions arose independently on the two continents. This important finding means that little, if any, of the mtDNA diversity that currently exists in Native Americans arose in Siberia prior to the Amerind migration. Therefore, the genetic diversity that exists on each continent can be considered proportional to the time that these populations have been separated.

Haplogroups C and D were the only haplogroups containing sufficient aboriginal Siberian and Native American mtDNA haplotypes to permit estimation of

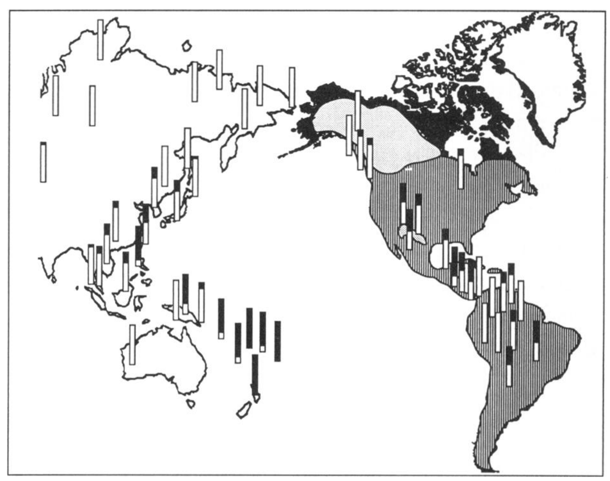


Figure 5 Distribution of the 9-bp COII-tRNA^{Lys} deletion around the Pacific Basin. Blackened portions of the bars indicate the frequency of the deletion. Data for Australia, Melanesia, and Polynesia are from Hertzberg et al. (1989) and Stoneking et al. (1990); for Japan (Japanese and Ainu), from Horai and Matsunaga (1986) and Harihara et al. (1992); for the Philippines, from Harihara et al. (1992); for the Altai region, from Shields et al. (1992); for Southeast Asia, from Ballinger et al. (1992); and for the Americas, from Torroni et al. (1992, 1993).

the continent-specific diversity. Haplogroup B was absent in Siberia, and Siberian haplogroup A was represented by only two complete haplotypes observed in the Evenks. The relationship between the Siberian and American branches of haplogroups C and D is shown in figure 6. Calculations of the sequence divergences of the Siberian branches of haplogroups C and D will be less accurate than those of the American haplogroups, because the Siberian data are based on the analysis of three populations, whereas the American data are derived from the analysis of 16 Amerind tribes. Consequently, the sampling error will be prone to make the Siberian values underestimates of mtDNA variation in Siberian populations.

With this limitation in mind, the sequence divergence values for the Siberian and American branches of haplogroups C and D were relatively consistent (table 6). In the Siberians, the group C divergence was 0.060% and the group D divergence was 0.040%, while in the Amerinds the value for group C was 0.096% and the value for group D was 0.053%. With a mtDNA evolution rate of 2%–4%/MYR (Stoneking et al. 1986; Cann et al. 1987; Wallace et al. 1987), the divergence time for group C mtDNAs would be 15,000–30,000 YBP in Siberians and 24,000-48,000 YBP in Amerinds, while the divergence time for group D would be 10,000–20,000

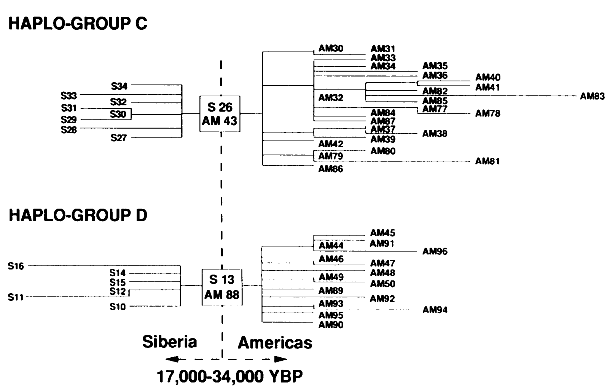


Figure 6 Radiation of haplogroups C and D in aboriginal Siberians (S) and Native Americans (AM). Amerind haplotypes are from Torroni et al. (1993). Group C haplotype S26 = AM43 and group D haplotype S13 = AM88 are the only mtDNA haplotypes shared between Siberians and Native Americans. The time of divergence was estimated from the sequence divergence data of table 6.

Table 6

Sequence Divergence and Radiation Time of Aboriginal Siberian and Amerind mtDNA Haplogroups

Haplogroup	*n*[a]	*N*[b]	Sequence Divergence (%)	Radiation Time[c] (years)
C:				
Siberian	9	51	.060	15,000–30,000
Amerind	23	61	.096	24,000–48,000
D:				
Siberian	7	21	.040	10,000–20,000
Amerind	16	60	.053	13,250–26,500
C + D:				
Siberian	16	72	.054	13,500–27,000
Amerind	39	121	.075	18,750–37,500
Combined C + D	53	193	.067	16,750–33,500

[a] No. of haplotypes.
[b] No. of subjects.
[c] Estimated using a mtDNA evolution rate of 2%–4%/MYR.

YBP in Siberians and 13,250–26,500 YBP in Amerinds. The average sequence divergence value for Siberian haplogroups C and D would be 0.054%, giving a divergence time of 13,500–27,000 YBP (table 6), while the corresponding values for the Amerinds would be 0.075% and 18,750–37,500 YBP.

Because the colonization of the Americas could not have occurred before that of Siberia, the best estimate for the colonization time of the Americas would be 18,750–37,500 YBP, with the migration into Siberia and then into the Americas occurring in relatively rapid succession. These data imply that the Americas were populated prior to the appearance of the Clovis lithic culture, for, even if the Siberian and Amerind values are averaged, the time required for the observed mtDNA differentiation would be 17,000–34,000 YBP (table 6 and fig. 6), a value in excess of the 13,500-YBP date generally accepted for the Clovis lithic culture.

Acknowledgments

We are indebted to Dr. J. V. Neel, Dr. M. D. Brown, and Ms. M. T. Lott for their intellectual contributions. This work was supported by NIH grant GM 46915 (to D.C.W.) and by NSF grant BSR 99101571 (to M.H.C.).

Appendix

Table A1

Polymorphic Restriction Sites Observed in 34 Distinct Siberian mtDNA Haplotypes (S1–S34)

Site[a]	Status by Haplotype[b]	Site[a]	Status by Haplotype[b]
	1111111111222222222233333 1234567890123456789012345678901234		1111111111222222222233333 1234567890123456789012345678901234
663e	0000000000000000011000000000000000	10180l	0000000000000010000000000000000000
853c/856a	1111011111111111111111111111111111	10394c	1111111111111110000011101111111111
1004o	1111111111111111111111111111110111	10397a	0000000111111110000011101111111111
1413l	1111111111111111111111111110111111	10656k	0000000000110000000000000000000000
1667c/1670a	1111111111111111111110111111111111	11900k	0001000000000000000000000000000000
1715c	1111111110111111111111111111000111	12026h/o	0000000000000001000000000000000000
1718e	0000000000000000000000000000000010	12170g/12171j	1111111111111111111111111111111011
3391e	0000000000000000000000001000000000	12406h/o	1111111111111111110111111111111111
3744e	0000000000000000000000100000000000	12629b	1111111111111111110111111111111111
4092e	0000000000000000000011000000000000	13259o/13262a	1111111111111111111111111000000000
4732k	0000000000000000000100000000000000	13702e[c]	0000000000000000000000000000000000
4769a[c]	0000000000000000000000000000000000	14168l	0100000000000000000000000000000000
4830n/4831f	0000000110000000000000000000000000	14199o[c]	0000000000000000000000000000000000
4877a	0000000000000001000000000000000000	14258m/14259j	1111111111111111111111101111111111
5176a	1111111110000000111111111111111111	14268g[c]	1111111111111111111111111111111111
5742i	1111111111111111101111111111111111	14368g[c]	0000000000000000000000000000000000
5971f	1111111111111101111111111111111111	15047e	1111111111111111111111101111111111
6915k	0000000000100000000000000000000000	15375g	1111011111111111111111111111111111
7025a[c]	1111111111111111111111111111111111	15606a	0000000000000000011000000000100000
7497e	1111111111111111111111111111111101	15883e	1111111111111111110111111111111111
7641a	1111101111111111111111111111111111	16208k	1111111111111111111110011111111110
7933j	1111111000000000000000000000000000	16303k	1111100111111111111011111111111111
8198a	0000000110000000000000000000000000	16310k	1111111111111111111011111111111111
8249b/8250e	0000000000000000000000000010000000	16389g/16390b	0000000000000000000000000001000000
8391e	0000000111111111111111111111111111	16398e	0000000010000000000000000000000000
8858f[c]	1111111111111111111111111111111111	16517e	1111111110100000000111010111111111
9751l/9753g	1111111111111111100111111111111111	4-bp insertion	0010000000000000000000000000000000
9820g	0000000000000000000011000000000000		

[a] Sites are numbered from the first nucleotide of the recognition sequence according to the published sequence (Anderson et al. 1981). The restriction enzymes used in the analysis are designated by the following single-letter code: a = *Alu*I; b = *Ava*II; c = *Dde*I; e = *Hae*III; f = *Hha*I; g = *Hin*fI; h = *Hpa*I; i = *Hpa*II; j = *Mbo*I; k = *Rsa*I; l = *Taq*I; m = *Bam*HI; n = *Hae*II; and o = *Hin*cII. Sites separated by a slash indicate either simultaneous site gains or site losses for two different enzymes or a site gain for one enzyme and a site loss for another because of a single inferred nucleotide substitution; these sites are considered to be only one restriction site polymorphism in the statistical analysis.

[b] A "1" indicates the presence of a site, and a "0" indicates the absence of a site, except for 4-bp insertions, where a "1" indicates the presence of a 4-bp insertion between the COII and $tRNA^{Lys}$ gene, and where a "0" indicates the absence of an insertion.

[c] Present or absent in all samples, contrary to the published sequence. The haplotype of the Senegalese used as an African outgroup was +9071, +2390j, −2758k, +3529h, +7025a, −7055a, +10394c, +10806g, +16517e.

References

Anderson S, Bankier AT, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139–152

Blanc H, Chen K-H, D'Amore MA, Wallace DC (1983) Amino acid change associated with the major polymorphic *Hin*cII site of Oriental and Caucasian mitochondrial DNAs. Am J Hum Genet 35:167–176

Bogoraz VG (1910) The Chukchee. Publications of the Jessup North Pacific Expedition. Vol 3, part 1. Leiden, New York

Brega A, Gardella R, Semino O, Morpurgo G, Astaldi Ricotti GB, Wallace DC, Santachiara Benerecetti AS (1986) Genetic studies on the Tharu population of Nepal: restriction endonuclease polymorphisms of mitochondrial DNA. Am J Hum Genet 39:502–512

Brown MD, Voljavec AS, Lott MT, Torroni A, Yang C-C, Wallace DC (1992) Mitochondrial DNA complex I and III mutations associated with Leber's hereditary optic neuropathy. Genetics 130:163–173

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cann RL, Wilson AC (1983) Length mutations in human mitochondrial DNA. Genetics 104:699–711

Chakraborty R, Weiss KM (1991) Genetic variation of the mitochondrial DNA genome in American Indians is at mutation-drift equilibrium. Am J Phys Anthropol 86:497–506

Crawford MH, Enciso VB (1982) Population structure of circumpolar groups of Siberia, Alaska, Canada and Greenland. In: Crawford MH, Mielke JH (eds) Ecology and population structure. Vol 2 in: Current developments in anthropological genetics. Plenum, New York, pp 51–91

Debets GF (1947) Sel'kups. Publication of the Institute of Ethnolography of the Soviet Academy of Sciences (new series). Vol 2. Moscow and Leningrad

Gershowitz H, Neel JV (1978) The immunoglobulin allotypes (Gm and Km) of twelve Indian tribes of Central and South America. Am J Phys Anthropol 49:289–301

Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. Proc Natl Acad Sci USA 77:6715–6719

Harihara S, Hirai M, Suutou Y, Shimizu K, Omoto K (1992) Frequency of a 9-bp deletion in the mitochondrial DNA among Asian populations. Hum Biol 64:161–166

Harihara S, Saitou N, Hirai M, Gojobori T, Park KS, Misawa S, Ellepola SB, et al (1988) Mitochondrial DNA polymorphism among five Asian populations. Am J Hum Genet 43:134–143

Hedges SB, Kumar S, Tamura K, Stoneking M (1991) Human origins and analysis of mitochondrial DNA sequences. Science 255:737–739

Hertzberg M, Mickleson KNP, Serjeantson SW, Prior JF, Trent RJ (1989) An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. Am J Hum Genet 44:504–510

Horai S, Gojobori T, Matsunaga E (1984) Mitochondrial DNA polymorphism in Japanese. Hum Genet 68:324–332

Horai S, Matsunaga E (1986) Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. Hum Genet 72:105–117

Jochelson W (1910) The Yukaghir and the Yukaghirized Tungus. Publications of the Jessup North Pacific Expedition. Vol 9, part 1. Leiden, New York

Karaphet TM, Sukernik RI, Osipova LP, Simchenko YB (1981) Blood groups, serum proteins, and red cell enzymes in the Nganasans (Tavghi): reindeer hunters from Taimir Peninsula. Am J Phys Anthropol 56:139–145

Kidd JR, Black FL, Weiss KM, Balazs I, Kidd KK (1991) Studies of three Amerindian populations using nuclear DNA polymorphisms. Hum Biol 63:775–794

Krauss ME (1988) Many tongues: ancient tales. In: Fitzhugh WW, Crowell A (eds) Crossroads of continents. Smithsonian Institution, Washington, DC, pp 145–150

Kreynovich EA (1978) Of some Yukagir-Uralic parallels. In: Sovetskoye finno ugrovedeniye. Vol 14 (in Russian). Academy of Sciences of the Estonian SSR, Tallin, pp 214–249

Levin MG, Potapov LP (eds) (1964) The peoples of Siberia. The University of Chicago Press, Chicago

Matsumoto H (1988) Characteristics of mongoloid and neighboring populations based on the genetic markers of human immunoglobulins. Hum Genet 80:207–218

Nei M, Tajima F (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. Genetics 105:207–217

Osipova LP, Sukernik RI (1983) Immunoglobulin allotypes in aboriginal populations of the Taimir populations. J Immunogenet 10:11–16

Posukh OL, Wiebe VP, Sukernik RI, Osipova LP, Karaphet TM, Schanfield MS (1990) Genetic studies of the Evens, an ancient human population of Eastern Siberia (USSR). Hum Biol 62:457–465

Rychkov YG, Sheremtyeva VA (1977) The genetic process in the system of ancient human isolates in North Asia. In: GA Harrison (ed) Population structure and human variation. Cambridge University Press, Cambridge, pp 47–108

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analyses for diagnoses of sickle cell anemia. Science 230:1350–1354

Schanfield MS (1980) Anthropological usefulness of highly polymorphic systems: HLA and immunoglobulin allotypes. In: Mielke JL, Crawford MH (eds) Current developments in anthropological genetics. Vol 1 in: Theory and methods. Plenum, New York, pp 65–85

——— (1992) Immunoglobulin allotypes (GM and KM) indicate multiple founding populations of Native Americans: evidence of at least four migrations to the New World. Hum Biol 64:381–402

Schurr TG, Ballinger SW, Gan Y-Y, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, et al (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613–623

Shields GF, Hecker K, Voevoda MI, Reed JK (1992) Absence of the Asian-specific region V mitochondrial marker in native Beringians. Am J Hum Genet 50:758–765

Shoffner JM, Lott MT, Lezza AMS, Seibel P, Ballinger SW, Wallace DC (1990) Myoclonic epilepsy and ragged-red fiber disease (MERRF) is associated with a mitochondrial DNA $tRNA^{Lys}$ mutation. Cell 61:931–937

Sokal RR, Rohlf FJ (1981) Taxonomic congruence in the Leptopodomorpha re-examined. Systematic Zool 30:309–325

Stoneking M, Bhatia K, Wilson AC (1986) Rate of sequence divergence estimated from restriction maps of mitochondrial DNAs from Papua New Guinea. Cold Spring Harb Symp Quant Biol 51:433–439

Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717–733

Stoneking M, Sherry ST, Redd AJ, Vigilant L (1992) New approaches to dating suggest a recent age for the human mtDNA ancestor. Philos Trans R Soc London [Biol] 337:167–175

Sukernik RI, Crawford MH, Osipova LP, Wiebe VP, Schanfield MS (1988) Peopling of the Americas in light of population genetic data. In: Tishkov VA (ed) Ecology of American Indians and the Eskimos (in Russian). Nauka, Moscow, pp 19–32

Sukernik RI, Lemza SV, Karaphet TM, Osipova LP (1981) Reindeer Chukchi and Siberian Eskimos: studies on blood groups, serum proteins, and red cell enzymes with regard to genetic heterogeneity. Am J Phys Anthropol 55:121–128

Sukernik RI, Osipova LP (1982) Gm and Km immunoglobulin allotypes in Reindeer Chukchi and Siberian Eskimos. Hum Genet 61:148–153

Sukernik RI, Osipova LP, Karaphet TM, Wiebe VP, Kirpichnokov GA (1986*a*) Genetic and ecological studies of aboriginal inhabitants of northeastern Siberia. I. GM haplotypes and their frequencies in ten Chukchi populations: genetic structure of Reindeer Chukchi (in Russian). Genetika 22:2361–2368

——— (1986*b*) Polymorphic blood systems, immunoglobulin allotypes, and other genetic markers in Asian Eskimos: genetic structure of the Bering Sea Eskimos (in Russian). Genetika 22:2369–2380

Sukernik RI, Osipova LP, Schanfield MS (1992) Distribution of GM allotypes and deleted IGHG1 haplotypes in the Sel'kups in comparison to the Forest Nentsi of northwestern Siberia. Clin Exp Immunogenet 9:15–23

Swofford D (1992) Phylogenetic analysis using parsimony (PAUP), version 3.0s. Illinois Natural History Survey, Champaign

Szathmary EJE (1981) Genetic markers in Siberian and Northern North American populations. Yearbook Phys Anthropol 24:37–73

——— (1984) Peopling of northern North America: clues from genetic studies. Acta Anthropol 8:79–109

Templeton AR (1991) Human origins and analysis of mitochondrial DNA sequences. Science 255:737

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, et al (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563–590

Torroni A, Schurr TG, Yang C-C, Szathmary EJE, Williams RC, Schanfield MS, Troup GA, et al (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. Genetics 130:153–162

Tugolukov VA (1979) Who are you, the Yukaghirs? (in Russian). Nauka, Moscow

——— (1985) Tungusic people (the Evenks and the Evens) of the middle and west Siberia (in Russian). Nauka, Moscow

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350–9354

Wallace DC, Garrison K, Knowler WC (1985) Dramatic founder effects in Amerindian mitochondrial DNAs. Amer J Phys Anthropol 68:149–155

Wallace DC, Torroni A (1992) American Indian prehistory as written in the mitochondrial DNA: a review. Hum Biol 64:403–416

Wallace DC, Ye J, Neckelmann SN, Singh G, Webster KA, Greenberg BD (1987) Sequence analysis of cDNAs for the human and bovine ATP synthase β subunit: mitochondrial genes sustain seventeen times more mutations. Curr Genet 12:81–90

Ward RH, Frazier BL, Dew-Jager K, Paabo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 88:1–5

Williams RC, Steinberg AG, Gershowitz H, Bennett PH, Knowler WC, Pettitt DJ, Butler W, et al (1985) GM allotypes in Native Americans: evidence for three distinct migrations across the Bering land bridge. Am J Phys Anthropol 66:1–19

Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC (1987) Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucleic Acids Res 15:529–542